AMENDMENTS TO THE CLAIMS

1. (Currently amended) A method, implemented by a computer, for encoding characters appearing in an area of an image in order to generate a corresponding output string of character codes, the method comprising the following steps, carried out by the computer:

identifying one or more sequences of the character codes that are ~~likely to be generated~~ known to result frequently due a segmentation error in application of a pattern recognition process, and associating a respective extension character code with each of the sequences, wherein each of the one or more sequences of the character codes is generated due to incorrect segmentation of a respective original character having a respective original character code;

dividing the area of the image into segments such that each segment contains approximately one character;

applying the pattern recognition process to each of the segments in order to generate an input string of character codes, the input string comprising a respective character code for each of the segments;

locating at least one of the sequences of the character codes in the input string, and replacing the at least one of the sequences with the respective extension character code so as to generate a modified string; and

determining the output string by ~~comparing~~ finding an approximate match between the modified string ~~to~~ and a known string in a directory of known strings by computing respective edit distances between the modified string and a plurality of the known strings based on respective costs of edit operations involving the extension character code, and selecting the one of the known strings responsively to the respective edit distances,

wherein a cost of zero is assigned to a transformation of the respective extension character code

associated with each of the sequences to the respective original character code.

2. (Original) The method according to claim 1, wherein the character codes that are generated by the pattern recognition process are selected from a predetermined set of eight-bit codes, and wherein associating the respective extension character code comprises assigning a respective eight-bit code that is not included in the predetermined set to replace each of the sequences.

3. (Original) The method according to claim 1, wherein applying the pattern recognition process comprises applying optical character recognition (OCR).

4-7. (Canceled)

8. (Currently amended) The method according to ~~claim 4~~ claim 1, wherein finding the approximate match comprises:

replacing each of the one or more sequences of the character codes in the known strings with the respective extension character code so as to create aliases that are respectively derived from the known strings;

adding the aliases to the directory; and

finding the approximate match between the modified string and one of the aliases, and

wherein outputting the one of the known strings comprises outputting the one of the known strings from which the one of the aliases is respectively derived.

9-24. (Canceled)